UNIVERSITY OF MICHIGAN

OpenClimate.org

The Validation of Climate Models:

The Development of Essential Practice

Richard B. Rood

University of Michigan

Wunderground.com

DCMIP, Boulder, 20120809

Deep Background

- As a manager at NASA
  - I felt a responsibility to deliver a series of model products addressing a specific set of scientific capabilities, on time, on budget.
  - I successfully argued that the modeling activity was a facility effort like an instrument.
  - As an instrument, I was required to provide a validation plan.
  - Many of my colleagues told me models could not be "validated."

Stubbornness

- I did not understand and accept that models could not be validated, though politics required me at times to talk about "evaluation."
  - It did not seem in any fundamental way to the validation of a satellite-based instrument.
- I thought a lot about validation and came to the conclusion that a validation strategy was critical to
  - Delivery on time
  - Delivery on budget
  - Ability to engage collaborators
  - Ability to communicate to customers
  - Credibility of the organization

But

- I have asked many of my colleagues whether or not climate models can be “validated” and they say “no.”
- But climate models in addition to have a presence in science have political and societal presences.
- What does it convey when scientists state that climate models cannot be validated?

Outline

- Introduction and Background
- Points of View on Validation
  - Philosophical
  - Computational Science
  - Software Engineering
- A Structured Validation Process
- A Set of Conclusions

Words of the Discussion

- Validation
- Verification
- Evaluation
- Testing

Widely accepted in Practice

- Calibration
- Certification
- Standardization
- Accreditation
- Trustworthiness

- The meanings of this words are nuanced by usage and audience.
- There are discipline-specific meanings of these words.
  - Philosophy
  - Science
  - Computational Science
  - Software engineering
  - …
- Audience
  - Scientist
  - Non-scientist

Background References

- [Oreskes et al., Science, 1994]
- Norton & Suppe, Changing Atmos., 2001
- [Guillemot, StudHistPhilModPhys, 2010]
- [Lenhard & Winsberg, StudHistPhilModPhys, 2010]
- [Post and Votta, PhysToday, 2005]
- [Michael et al. IEEE Software, 2011]
- [Farber, Berkeley Law, 2007]
- [Science Integrity: Climate Models: 1995]
  - (1200 pages, 55MB, Congressional Testimony)

Outline

- Introduction and Background
- Points of View on Validation
  - – Philosophical
  - – Computational Science
  - – Software Engineering
- A Structured Validation Process
- A Set of Conclusions

Verification and Validation

- Philosophy, Science, (etymology)
  – Verification – establishment of truth
  – Validation – strong / supported by authority
- Computational science
  – Verification – code works correctly
  – Validation – model captures essential physical phenomena
- Software engineering
  – Verification – code built correctly
  – Validation – code meets requirements of design

Climate modeling belongs to all of these domains

Validation

- American Heritage Dictionary
  - To declare or make legally valid
  - To mark with an indication of official sanction
  - To establish the soundness of: corroborate
- Valid →
  - Well grounded; just
  - Producing the desire results; efficacious
  - Having legal force; effective or binding
  - Containing premises from which the conclusion may be logically derived (logic)
  - Correctly inferred or deduced from a premise (logic)

A thread of arguments

- Oreskes et al. → Models cannot be verified or validated
  - Open systems
  - Underdetermination, non-uniqueness
- Norton and Suppe → Models are pervasive in all forms of science → Uniqueness of solution and single "truth" is a false pursuit → If models cannot be validated, then science is unfounded as a way to generate knowledge → absurdity
  - No real difference between numerical and experimental science
  - Role of theory, data and geophysics
  - Uniqueness is not a measure of validity
- Guillemot and other studies → Describe practice of model evaluation → Models lead to conclusions that can be evaluated and, *de facto*, validated.
- Concept of Pluralism and Community-based evaluation

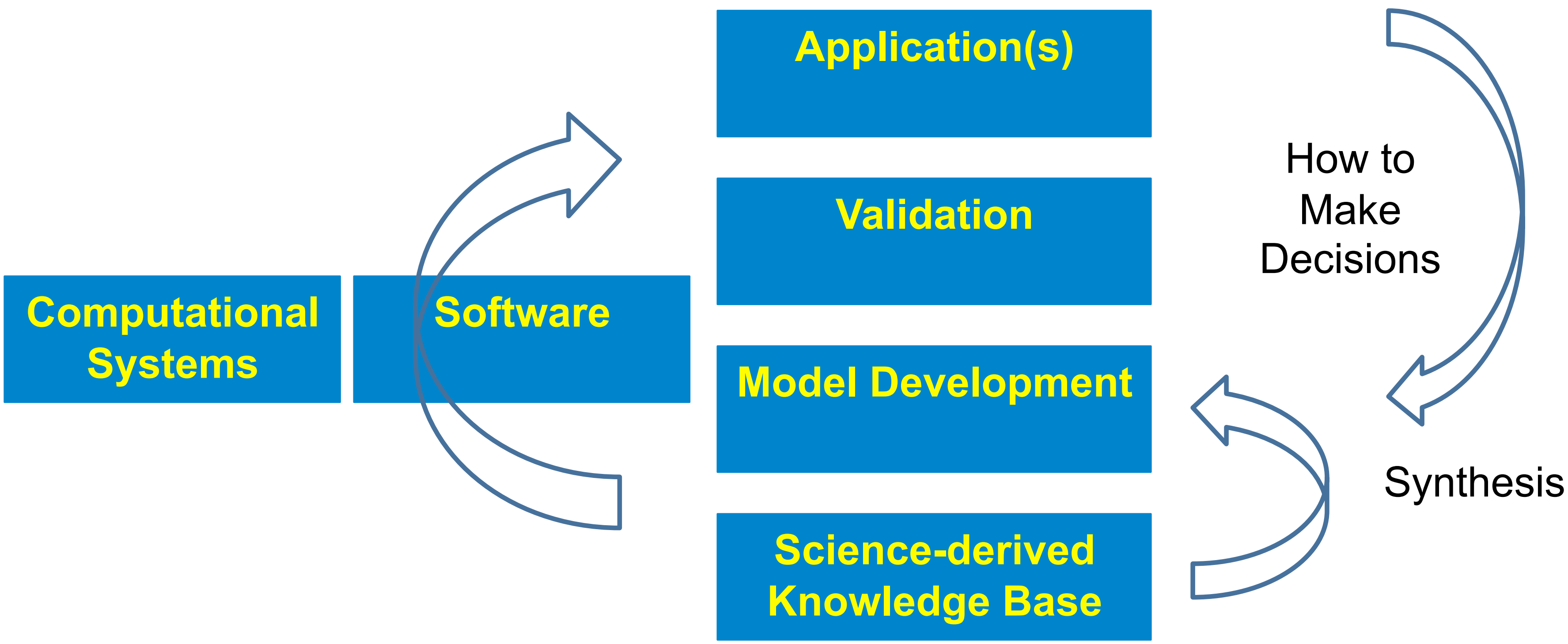
Continued thread of arguments

- Post et al.: Computational science is a new "kind" of science that requires verification and validation → verification and validation are underrepresented in the enterprise as a whole
- Validation contributes to trustworthiness
- Going forward: Evaluation of models can be described and codified to establish a validation plan to support model application and knowledge generation (my premise)

Outline

- Introduction and Background
- Points of View on Validation
  - Philosophical
  - Computational Science
  - Software Engineering
- A Structured Validation Process
- A Set of Conclusions

Functions in Model Development



Application(s) and Validation

Application(s)

Application: Why is the model being built?

Validation

Validation: Is the model addressing its goal?

Model building is an integrating or synthesizing process. The identification of the model application(s) provides the primary framing for what to choose out of the body of science-based knowledge. The development of a validation plan provides a way to evaluate whether or not the model is addressing the application. The validation process further defines decision making, and it links vision and goals to implementation.

In addition to integrating and synthesizing science-based decisions, integration and synthesis is required across the Computational Systems and Software.

The “model as a whole” needs to managed.

Validation

Validation

Validation is an essential part of the scientific method. We regularly practice validation with comparisons of simulation to observations, with comparisons of multiple methods to address the same problem, with peer review, with the practice of independent researchers obtaining the same result.

<u>What does this imply for climate modeling?</u>

- The need for organizational design of a validation plan to evaluate the performance of the entire system's ability to address the application(s); testable design criteria.
- The need for the organization to develop of an "independent" validation process.
- The need to document the validation plan and validation process prior to development cycle.

Thinking about Validation Structure

Computational Systems	Software	Computational Science	Natural Science

- Need to think about robust strategies for testing and evaluation in all of venues
- Need to understand if you are a natural scientist, then you are in a system with these other disciplines and professions
- Testing and evaluation in one part of the system is NOT independent of or irrelevant to the other parts of the system

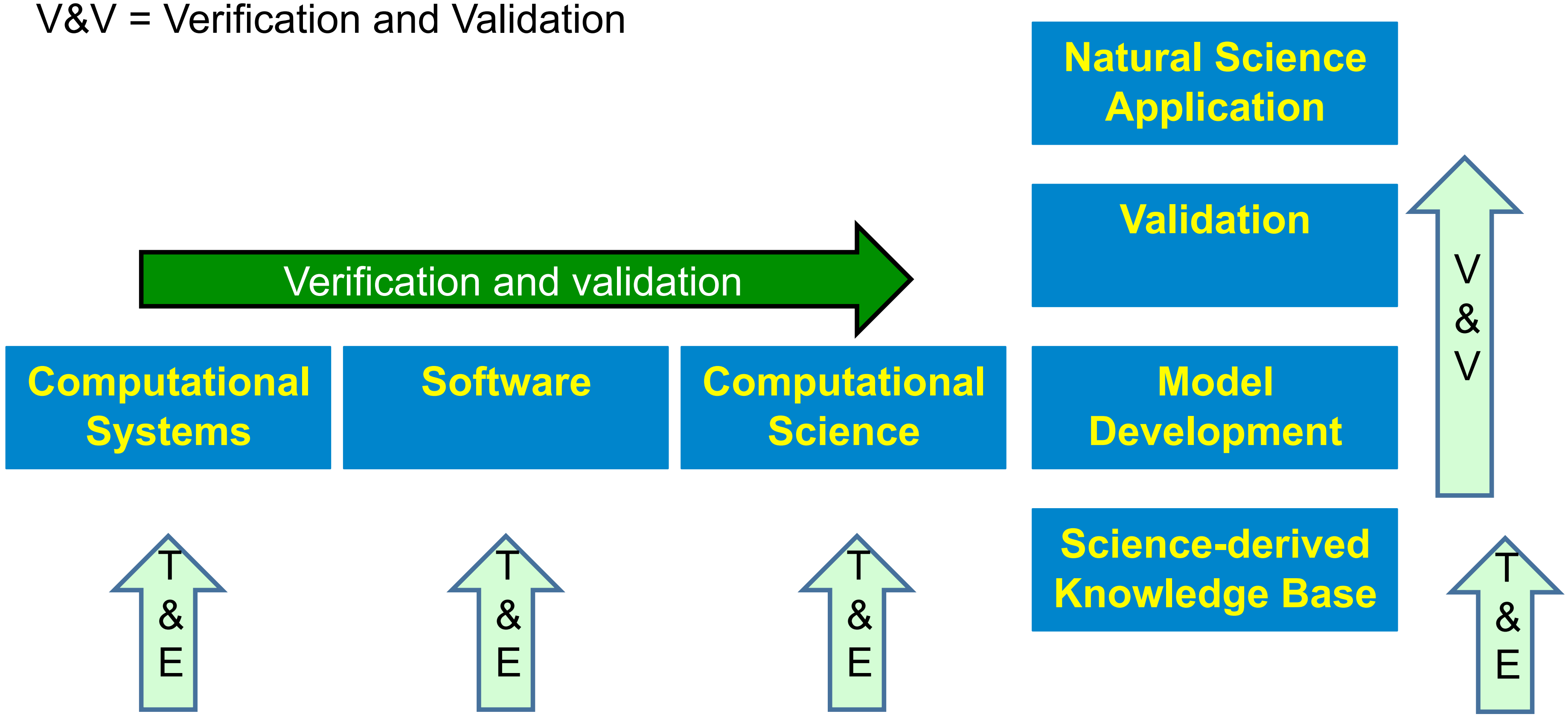
Software and Science: Testing

[Clune and Rood: Test Driven Development]

- Software testing can
  - improve scientific code
  - decrease development time
- From a software expert's perspective
  - Systems Testing
  - Integration Testing
  - Unit Testing
- Requires developing test plan (in advance)

Integrated Testing and Evaluation Required Throughout

T= Testing
E = Evaluation
V&V = Verification and Validation



Evaluation Practice

Evaluation Practice

- Integrated quantities
- Phenomenological comparison
- Prediction
- Correlated physics
- Processes
- Heuristic
- Theory

At an Institution

- Most or all of these practices are present at different phases of model development and implementation
- Often dependent on interests and expertise of individuals
- Institutional and community conventions evolve

Elements of Validation

- Monitoring & Quality Assessment
- Component Validation
- Initialized Forecasts
- Systems Validation
- Scientific Validation

Quantification and Automation

DCMIP workshop

The Hard Part

Open ended

[DAO Algorithm Theoretical Basis Document, NASA, 1996]

Important to distinguish between and to manage the interfaces of testing, verification, and validation of software practice and science model implementation.

Thinking about elements of validation

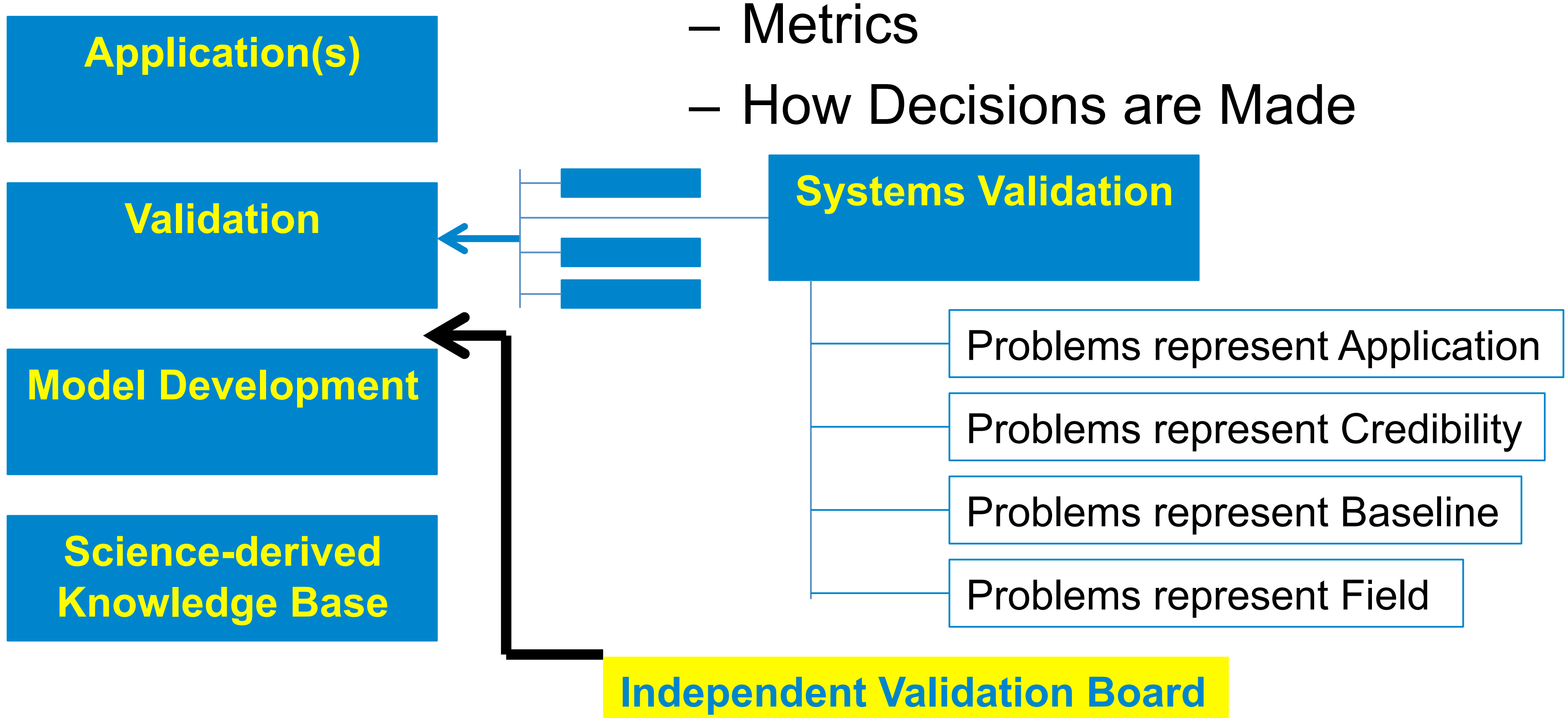
- [Lenhard & Winsberg, StudHistPhilModPhys, 2010]()
  - Focus on complexity of models and Earth's climate
  - There is little evidence to support the idea that we can link climate model performance to strengths and weaknesses of component models
  - A process that combines knowledge from observations, theory (analytics), simulation
  - Can we challenge this evidence?
  - Can we define tests and metrics in advance?
  - How do we know when we have succeeded?

Requirements for Validation Plan

- Application / Purpose for Model Development
  - Decision making
  - Focus on Integrated Model to Address Application rather than the Model Components
  - Focus of Model, Software, and Computational Systems
- Multiple Sources of Evaluation Information
  - Observations
  - Consistency
- Independent Validation Scientists
- Process to Support Validation
  - Documentation
  - Metrics
  - How Decisions are Made
  - …

Requirements for Validation Plan

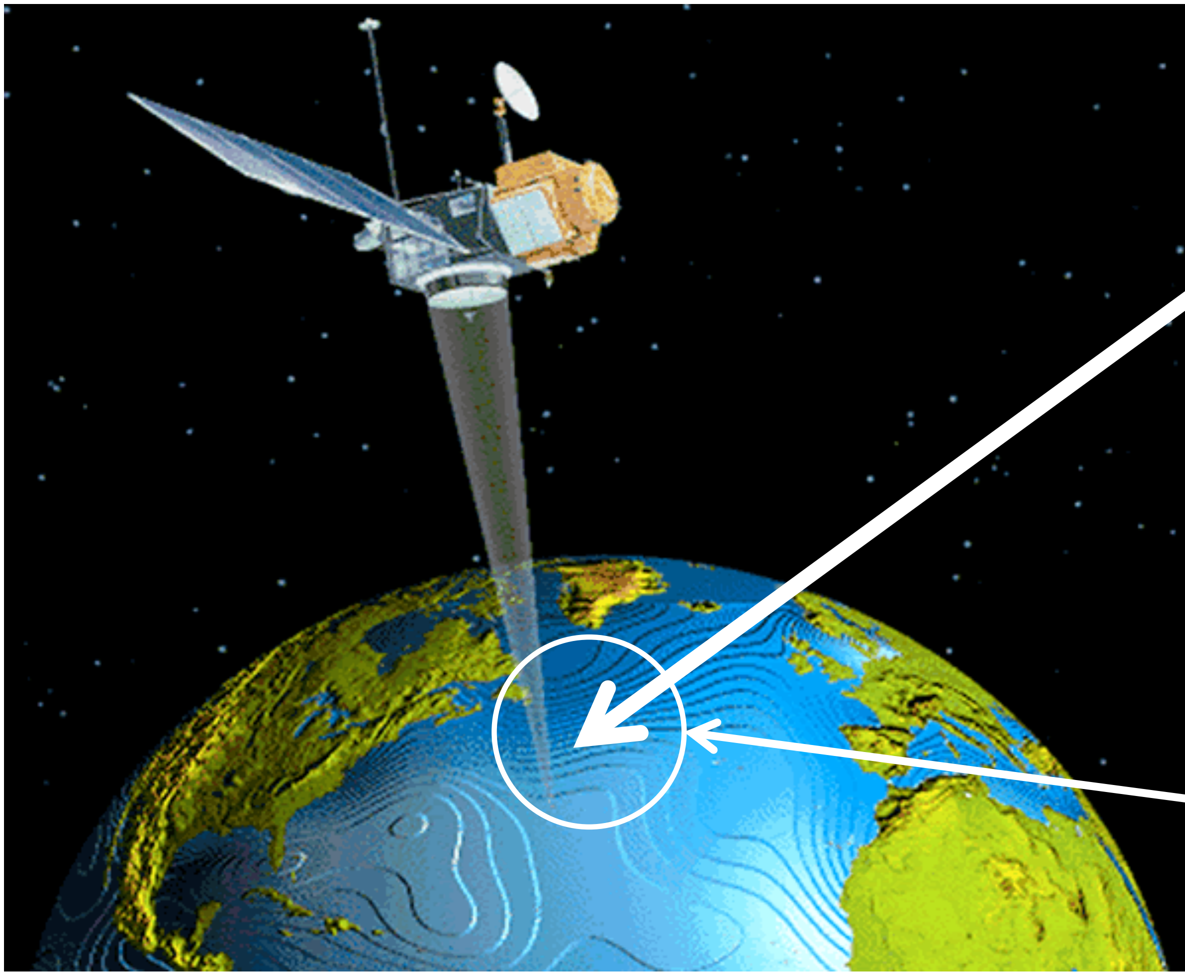
- Process to Support Validation
  - Documentation
  - Metrics
  - How Decisions are Made



UNIVERSITY OF MICHIGAN

OpenClimate.org

An Important Attribute of Climate Model Validation (a NASA-based example)

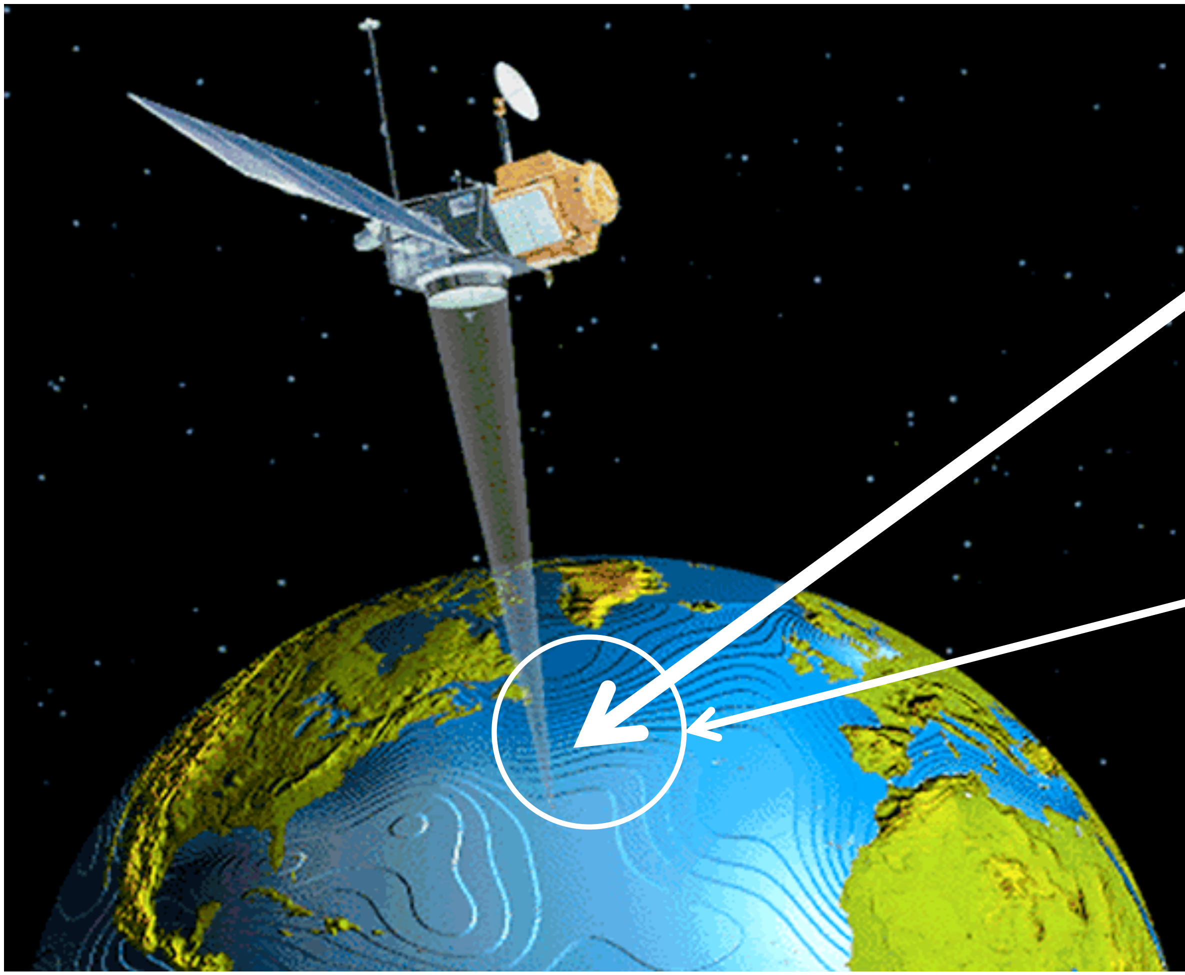


Independent Observations

- Planes
- Ships
- Balloons
- Buoys
- Weather Station

Map in space and time and "validate."

An Important Attribute of Climate Model Validation (a NASA-based example)



Independent Observations

- Planes
- Ships
- Balloons
- Buoys
- Weather Station

Map in space and time and “validate.”

In this validation attention is reduced to a “single” focus, a number.

- **Model validation focuses on ever expanding complexity.**

Outline

- Introduction and Background
- Points of View on Validation
  - Philosophical
  - Computational Science
  - Software Engineering
- A Structured Validation Process
- A Set of Conclusions

Some of my Initial Claims

- Delivery on time
  - Stops development from running open loop
- Delivery on budget
  - Limits scope of effort, and maps, directly, computational resources to development
- Ability to engage collaborators
  - Collaborators know what they are working towards
- Ability to communicate to customers
  - Performance metrics for specific problems
  - Documented process for non-scientist users
- Credibility of the organization
  - Provide products on time and on budget
  - Scientific method defines organizational goals
    - As contrasted with an organization of scientists

Some criticisms

- Climate models can't be validated
- Would hurt "the science"
  - Removes critical resources
  - Hands validation to non-scientists
  - Prevents latest science from getting into the system
- Requires overhead of management and governance that:
  - Removes critical resources
  - Takes too much time
  - Removes valuable trained scientists
  - Hands decision making to non-experts
  - Is contrary to "science"
- Hurts creativity, stifles innovation
  - Discoveries and breakthroughs come from unexpected places

Reasons to Formalize Practice

- Basic credibility of the field
  - Scientific
  - Broader applications
  - Baseline to measure progress
  - Baseline to describe uncertainty
- Improve our ability to communicate
- Improve organizations ability to deliver on schedule and on budget → Fundamentally strategic and aids implementation.
  - Improve ability to define and utilize resources
- Improve the ability to incorporate a community of researchers into the field
- Organizations that adhere to the scientific method
  - Rather than an organization full of science-minded scientists

ESSENTIAL PRACTICE

More Rood-like References

- DAO Algorithm Theoretical Basis Document, NASA, 1996
- UoM Class References: Model Validation
- Steve Easterbrook: Serendipity
- Rood Blog Data Base
  - Validation
- Lemos and Rood: Uncertainty
- Clune and Rood: Test Driven Development

UNIVERSITY OF MICHIGAN

OpenClimate.org

Background References

- Oreskes et al., Science, 1994
- Norton & Suppe, Changing Atmos., 2001
- Guillemot, StudHistPhilModPhys, 2010
- Lenhard & Winsberg, StudHistPhilModPhys, 2010
- Post and Votta, PhysToday, 2005
- Michael et al. IEEE Software, 2011
- Farber, Berkeley Law, 2007
- Science Integrity: Climate Models: 1995
  - (1200 pages, 55MB, Congressional Testimony)